IN THE SPECIFICATION

Please amend the specification by replacing the paragraph on page 25 lines 8 through 20 with the following Table A and paragraph:

Table A: Codon degeneracy of amino acids

Amino acid	One letter	Three letter	Codons
Alanine	A	Ala	GCA GCC GCG GCT
Cysteine	C	Cys	TGC TGT
Aspartic acid	D	Asp	GAC GAT
Glutamic acid	E	Glu	GAA GAG
Phenylalanine	F	Phe	TTC TTT
Glycine	G	Gly	GGA GGC GGG GGT
Histidine	H	His	CAC CAT
Isoleucine	I	Ile	ATA ATC ATT
Lysine	K	Lys	AAA AAG
Leucine	L	Leu	TTA TTG CTA CTC CTG CTT
Methionine	M	Met	ATG
Asparagine	N	Asn	AAC AAT
Proline	P	Pro	CCA CCC CCG CCT
Glutamine	Q	Gln	CAA CAG
Arginine	R	Arg	AGA AGG CGA CGC CGG CGT
Serine	S	Ser	AGC AGT TCA TCC TCG TCT

Threonine	T	Thr	ACA ACC ACG ACT
Valine	V	Val	GTA GTC GTG GTT
Tryptophan	W	Trp	TGG
Tyrosine	Y	Tyr	TAC TAT

Certain amino acids may be substituted for other amino acids in a protein sequence without appreciable loss of the desired activity. It is thus contemplated that various changes may be made in the peptide sequences of the disclosed protein sequences, or their corresponding nucleic acid sequences without appreciable loss of the biological activity.

Please replace the paragraph in the specification at page 39 lines 10 through 17 with the following amended paragraph:

The short nucleic acid sequences may be used as probes and specifically as PCR probes. A PCR probe is a nucleic acid molecule capable of initiating a polymerase activity while in a double-stranded structure with another nucleic acid. Various methods for determining the structure of PCR probes and PCR techniques exist in the art. Computer generated searches using programs such as Primer3 (www[[.]]-genome.wi.mit.edu/cgi-bin/primer/primer2.cgi), STSPipeline (www-genome.wi.mit.edu/cgi-bin/www.STS_Pipeline), or GeneUp (Pesole, et al., BioTechniques 25:112-123, 1998), for example, can be used to identify potential PCR primers.

Please replace the paragraph in the specification at page 78 lines 7 through 16 with the following amended paragraph:

PHRED is used to call the bases from the sequence trace files (~~http://~~www[[.]]-mbt.washington.edu). Phred uses fourier methods to examine the four base traces in the region surrounding each point in the data set in order to predict a series of evenly spaced predicted locations. That is, it determines where the peaks would be centered if there were no

compressions, dropouts, or other factors shifting the peaks from their "true" locations. Next, PHRED examines each trace to find the centers of the actual, or observed peaks and the areas of these peaks relative to their neighbors. The peaks are detected independently along each of the four traces so many peaks overlap. A dynamic programming algorithm is used to match the observed peaks detected in the second step with the predicted peak locations found in the first step.

Please replace the two paragraphs beginning "Example 3:" in the specification at page 79 line 10 through page 81 line 5 with the following two amended paragraphs:

Example 3: Identifying genes within a genomic BAC library

This example illustrates the identification of combigenes within the rice genomic contig library as assembled in Example 2. The genes and partial genes that are embedded in such contigs are identified through a series of informatic analyses. The tools to define genes fall into two categories: homology-based and predictive-based methods. Homology-based searches (*e.g.*, GAP2, BLASTX supplemented by NAP and TBLASTX) detect conserved sequences during comparisons of DNA sequences or hypothetically translated protein sequences to public and/or proprietary DNA and protein databases. Existence of an *Oryza sativa* gene is inferred if significant sequence similarity extends over the majority of the target gene. Since homology-based methods may overlook genes unique to *Oryza sativa*, for which homologous nucleic acid molecules have not yet been identified in databases, gene prediction programs are also used. Predictive methods employed in the definition of the *Oryza sativa* genes included the use of the GenScan gene predictive software program which is available from Stanford University (*e.g.*, at the website: ~~gnomic/stanford.edu/GENSCANW.html~~www-gnomic/stanford.edu/GENSCANW.html, and the Genemark.hmm for Eukaryotes program from Gene Probe, Inc (Atlanta, GA) www-geneprobe.net/index.htm~~www.geneprobe.net/index.htm~~). GenScan, in general terms, infers the presence and extent of a gene through a search for "gene-like" grammar. GeneMark.hmm searches a file containing DNA sequence data for genes. It

employs a Hidden Markov Model algorithm with a species-specific inhomogeneous Markov model of gene-encoding regions of DNA.

The homology-based methods that are used to define the *Oryza sativa* gene set included GAP2, BLASTX supplemented by NAP and TBLASTX. For a description of BLASTX and TBLASTX *see* Coulson, *Trends in Biotechnology* 12:76-80 (1994) and Birren *et al., Genome Analysis,* 1:543-559 (1997). GAP2 and NAP are part of the Analysis and Annotation Tool (AAT) for Finding Genes in Genomic Sequences which was developed by Xiaoqiu Huang at Michigan Tech University [[ ]]and is available at the web site www-genome.cs.mtu.edu/~~genome.cs.mtu.edu/~~. The AAT package includes two sets of programs, one set DPS/NAP (referred to as "NAP") for comparing the query sequence with a protein database, and the other set DDS/GAP2 (referred to as "GAP2") for comparing the query sequence with a cDNA database. Each set contains a fast database search program and a rigorous alignment program. The database search program identifies regions of the query sequence that are similar to a database sequence. Then the alignment program constructs an optimal alignment for each region and the database sequence. The alignment program also reports the coordinates of exons in the query sequence. *See* Huang, *et al., Genomics 46*: 37-45 (1997). The GAP2 program computes an optimal global alignment of a genomic sequence and a cDNA sequence without penalizing terminal gaps. A long gap in the cDNA sequence is given a constant penalty. The DNA-DNA alignment by GAP2 adjusts penalties to accommodate introns. The GAP2 program makes use of splice site consensuses in alignment computation. GAP2 delivers the alignment in linear space, so long sequences can be aligned. *See* Huang, *Computer Applications in the Biosciences 10* 227-235 (1994). The GAP2 program aligns the *Oryza sativa* contigs with a library of 42,260 *Oryza sativa* cDNAs.

Please replace the paragraph at page 81 lines 13 through 17 with the following amended paragraph:

NAP takes a nucleotide sequence, translates it in three forward reading frames and three reverse complement reading frames, and then compares the six translations against a protein sequence database (*e.g.* the non-redundant protein (*i.e.*, nr-aa database maintained by the National Center for Biotechnology Information as part of GenBank and available at the web site: www-ncbi.nlm.nih.gov~~www.ncbi.nlm.nih.gov~~).

Please replace the paragraph at page 86 lines 3 through 7 with the following amended paragraph:

Putative promoter sequences are also searched with matrices for the TATA box, GC box (factor name: V_GC_01) and CCAAT box (factor name: F_HAP234_01). The matrix for the TATA box is from the Eukaryotic Promoter Database (www-epd.isb-sib.ch/**~~http://www.epd.isb-sib.ch/~~**) and the matrices for the GC box and the CCAAT box are from Transfac (www-transfac.gbf.de/TRANSFAC/**~~http://transfac.gbf.de/TRANSFAC/~~**).

Please replace the paragraph beginning "Example 5" at page 87 line 22 through page 88 line 10 with the following amended paragraph:

Example 5: Identifying promoters in the genomic BAC library using an expression assay

Promoters may also be identified based on quantitative analysis of genes that are cis-associated with candidate promoters, (i.e. the native genes). In this method, the native genes associated with SEQ ID NO:1 through SEQ ID NO:57,467 are analyzed on a digital northern blot. Digital northern data can be generated from EST sequencing, SAGE and other methods, which in effect count RNA molecules expressed in cell. This data can be generated as needed, or is generally available to the public on a number of web sites (*e.g.*, www-tigr.org~~www.tigr.org~~). Data can be obtained from any plant species, although data on rice gene expression is particularly preferred. Promoters are selected based on the expression information of the digital northern. For [[]]example, identifying genes expressing genes under stress-related conditions

would provide a group of promoters able to confer such stress-inducible expression to other genes.

Please amend the specification on page 91 line 13 through page 93 line 20 by replacing the section beginning "Table 2" with the following:

Table 2

Transcription Factor Name	Sequence Motif name	Sequence Motif	Sequence Motif Length	Maximum mismatches allowed	Reference for transcription factors and sequence motifs
Fac006	ABADESI1	RTACGTGGCR	10	1	PLACE
Fac007	ABADESI2	GGACGCGTGGC	11	2	PLACE
Fac010	ABFOS	GCATCTTTACTTTAGCATC	19	6	PLACE
Fac016	ABRE3OSRAB16	GTACGTGGCGC	11	2	PLACE
Fac016	ABREATRD22	RYACGTGGYR	10	0	PLACE
Fac020	ABREOSRAB21	ACGTSSSC	8	0	PLACE
Fac021	ABREOSRGA1	CCACGTGG	8	0	PLACE
Fac021	ABRETAEM	GGACACGTGGC	11	2	PLACE
Fac022	ACGTABOX	TACGTA	6	0	PLACE
Fac031	AMYBOX1	TAACARA	7	0	PLACE
Fac032	AMYBOX2	TATCCAT	7	0	PLACE
Fac060	DREDR1ATRD29AB	TACCGACAT	9	1	PLACE
Fac064	EREGCCNTCHN	TAAGAGCCGCC	11	2	PLACE

Transcription Factor Name	Sequence Motif name	Sequence Motif	Sequence Motif Length	Maximum mismatches allowed	Reference for transcription factors and sequence motifs
Fac066	GARE2R	TAACARANTCYGG	14	2	PLACE
Fac068	GBOXRELOSAMY3	CTACGTGGCCA	11	2	PLACE
Fac070	GLUTAACAOS	AACAAACTCTAT	12	2	PLACE
Fac071	1OSGT2GLUTEBOX	ATATCATGAGTCACTTCA	18	4	PLACE
Fac071	1OSGT2GLUTEBOX	ATATCATGAGTCACTTCA	18	4	PLACE
Fac071	1OSGT3GLUTEBOX	TATCTAGTGAGTCACTTCA	19	5	PLACE
Fac071	1OSGT3GLUTEBOX	TATCTAGTGAGTCACTTCA	19	5	PLACE
Fac072	2OSGT2GLUTEBOX	TCCGTGTACCA	11	2	PLACE
Fac072	2OSGT2GLUTEBOX	TCCGTGTACCA	11	2	PLACE
Fac072	2OSGT3GLUTEBOX	CTTTTGTGTACCTTA	15	3	PLACE
Fac072	2OSGT3GLUTEBOX	CTTTTGTGTACCTTA	15	3	PLACE
Fac073	GLUTEBP1OS	AAGCAACACACAAC	14	3	PLACE
Fac074	GLUTEBP2OS	ATGCTCAATAGATATAAGT	19	5	PLACE
Fac075	GLUTECOREOS	CTTTCGTGTAC	11	2	PLACE
Fac079	GT2OSPHY	AGCGGTAATT	9	1	PLACE
Fac105	MYBGAHV	TAACAAA	7	0	PLACE

Transcription Factor Name	Sequence Motif name	Sequence Motif	Sequence Motif Length	Maximum mismatches allowed	Reference for transcription factors and sequence motifs
Fac129	PROLAMINBOX	CACATGTGTAAAGGT	15	4	PLACE
Fac135	RGATAOS	CAGAAGATA	9	1	PLACE
Fac136	RNFG1OS	GATCATCGATC	11	2	PLACE
Fac137	RNFG2OS	CCAGTGTGCCCCTGG	15	4	PLACE
Fac139	RYREPEAT4	TCCATGCATGCAC	13	3	PLACE
Fac139	RYREPEAT4	TCCATGCATGCAC	13	3	PLACE
Fac139	RYREPEATGMGY2	CATGCAT	7	0	PLACE
Fac139	RYREPEATGMGY2	CATGCAT	7	0	PLACE
Fac139	RYREPEATLEGUMINBOX	CATGCAY	7	0	PLACE
Fac139	RYREPEATLEGUMINBOX	CATGCAY	7	0	PLACE
Fac139	RYREPEATVFLEB4	CATGCATG	8	0	PLACE
Fac139	RYREPEATVFLEB4	CATGCATG	8	0	PLACE
Fac149	SITEIIAOSPCNA	TGGGCCCGT	9	1	PLACE
Fac150	SITEIIBOSPCNA	TGGTCCCAC	9	1	PLACE
Fac151	SITEIOSPCNA	CCAGGTGG	8	1	PLACE
Fac163	AACAOSGLUB1	CAACAAACTATATC	14	3.5	PLACE

Transcription Factor Name	Sequence Motif name	Sequence Motif	Sequence Motif Length	Maximum mismatches allowed	Reference for transcription factors and sequence motifs
Fac165	ACGTOSGLUB1	GTACGTG	7	0	PLACE
Fac180	GT1CONSENSUS	GRWAAW	6	0	PLACE
Fac201	PYRIMIDINEBOXOSRAMY1A	CCTTTT	6	0	PLACE
Fac218	ABREMOTIFAOSOSEM	TACGTGTC	8	0.5	PLACE
Fac219	ABREMOTIFIIIOSRAB16B	GCCGCGTGGC	10	1.5	PLACE
Fac220	ABREMOTIFIOSRAB16B	AGTACGTGGC	10	1.5	PLACE
Fac223	CE3OSOSEM	AACGCGTGTC	10	1.5	PLACE
Fac267	POLASIG2	AATTAAA	7	0	PLACE
OS__A-box	OS__A-box	TATCCATCCATCC	13	3	PlantCARE
OS__A-box2	OS__A-box2	AATAACAAACTCC	13	3	PlantCARE
OS__AACA	OS__AACA	TAACAAACTCCA	12	2.5	PlantCARE
OS__ABRE	OS__ABRE	GACACGTACGT	11	2	PlantCARE
OS__ABRE2	OS__ABRE2	ACGTACGTGTCGCGC	15	4	PlantCARE
OS__AP-2-like	OS__AP-2-like	CGCGCCGG	8	0.5	PlantCARE
OS__AP-2-like2	OS__AP-2-like2	CGACCAGG	8	0.5	PlantCARE
OS__ATGCAAAT	OS__ATGCAAAT	ATACAAAT	8	0.5	PlantCARE

Transcription Factor Name	Sequence Motif name	Sequence Motif	Sequence Motif Length	Maximum mismatches allowed	Reference for transcription factors and sequence motifs
OS__CE3	OS__CE3	GACGCGTGTC	10	1.5	PlantCARE
OS__GATT	OS__GATT	CTCCTGATTGGA	12	2.5	PlantCARE
OS__GCN4	OS__GCN4	TGWGTCA	7	0	PlantCARE
OS__GCN4_2	OS__GCN4_2	CAAGCCA	7	0	PlantCARE
OS__P-box	OS__P-box	GCCTTTTGAGT	11	2	PlantCARE
OS__P-box2	OS__P-box2	CCTTTTG	7	0	PlantCARE
OS__Prolamin_box	OS__Prolamin_box	TGCAAAGT	8	0.5	PlantCARE
OS__Skn-1	OS__Skn-1	GTCAT	5	0	PlantCARE
OS__TATC-box	OS__TATC-box	TATCCCA	7	0	PlantCARE
OS__TGGCA	OS__TGGCA	GACACCAAGTGGCA	14	3.5	PlantCARE
OS__light	OS__light	AACCAATCTCATCCATCC	18	5.5	PlantCARE
AS_RF2A_01	AS_RF2A_01	CCAGTGTGGCGCTGG	15	4	TRANS
AT_RS1A_01	AT_RS1A_01	CTTCCACGTGGCA	13	3	TRANS
PV_GRP18_01	PV_GRP18_01	TGGATGTGGAAGACAGCA	18	5.5	TRANS
RICE_ACT_01	RICE_ACT_01	GCCCAACCCAACCCAAC	17	5	TRANS
RICE_AGB_01	RICE_AGB_01	GCCACGTAAG	10	1.5	TRANS

Transcription Factor Name	Sequence Motif name	Sequence Motif	Sequence Motif Length	Maximum mismatches allowed	Reference for transcription factors and sequence motifs
RICE_AGB_03	RICE_AGB_03	GCCACGTCAG	10	1.5	TRANS
RICE_EM_01	RICE_EM_01	TACGTGT	7	0	TRANS
RICE_EM_02	RICE_EM_02	GACGTGT	7	0	TRANS
RICE_GL51_01	RICE_GL51_01	AAGTCATAACTG	12	2.5	TRANS
RICE_GL51_02	RICE_GL51_02	CCATGTCATATT	12	2.5	TRANS
RICE_GL51_03	RICE_GL51_03	AATGATGTGTCAAT	14	3.5	TRANS
RICE_GL51_04	RICE_GL51_04	TTCCGTGTACCAC	13	3	TRANS
RICE_GL51_05	RICE_GL51_05	TGAGTCA	7	0	TRANS
RICE_GLU2_01	RICE_GLU2_01	CCTTTCGTGTACC	13	3	TRANS
RICE_GLUB1_01	RICE_GLUB1_01	CTGAGTCAT	9	1	TRANS
RICE_NITR_01	RICE_NITR_01	CACGTCAC	8	0.5	TRANS
RICE_RAB16A_01	RICE_RAB16A_01	TACGTGGCNNNNCCGC CGCGCCT	23	6	TRANS
RICE_RAB16A_03	RICE_RAB16A_03	GTACGTGG	8	0.5	TRANS
TAF-1AS_	TAF1_01	GCAACGTGGC	10	1.5	TRANS
TAF-1RICE_	RAB16B_01	GGTACGTGGCG	11	2	TRANS

Transcription Factor Name	Sequence Motif name	Sequence Motif	Sequence Motif Length	Maximum mismatches allowed	Reference for transcription factors and sequence motifs
WHEAT_H3_01	WHEAT_H3_01	CCACGTCA	8	0.5	TRANS
Seed_sp	Seed_AACA_motif	AACAAACTCTATC	13	3	lit1
Seed_sp	Seed_GCN4	GTGAGTCAC	9	1	lit1
SugRep	ACGTABOX	TACGTA	6	0	lit2
SugRep	TCmotif	TATCCAY	7	0	lit2
Amy3	DAMYBOX2	TATCCAT	7	0	lit3
Amy3	DGBOXRELOSAMY3	CTACGTGGCCA	11	2	lit3

***Column Headings for Table 2**

Please replace the paragraph at page 94 lines 2 through 8 with the following amended paragraph:

Reference for transcription factors and sequence motifs: Motifs and transcription factors are found in one of three databases: PLACE, PlantCARE or TRANS (respectively, www-dna.affrc.go.jp/htdocs/PLACE/ ~~**http://www.dna.affrc.go.jp/htdocs/PLACE/**~~, www-sphinx.rug.ac.be:8080/PlantCARE/index.htm~~**http://sphinx.rug.ac.be:8080/PlantCARE/index.htm**~~, www-transfac.gbf.de/TRANSFAC/, or~~**orhttp://transfac.gbf.de/TRANSFAC/or**~~ Yoshihara et al., FEBS Letters 383, 1996, pp 213-218; or Toyofuku K et al. FEBS Lett 428:275-280 (1998) or lit3 (Huang et al Plant Mol Biol 14:655-668 (1990)).